

September 1, 1959

CONTRIBUTIONS TO THE THEORY OF RANK ORDER STATISTICS:

Computation Rules for Probabilities of Rank Orders

I. Richard Savage 1/

Technical Report No. 9

University of Minnesota

Minneapolis, Minnesota

Reproduction in whole or in part is permitted for any purpose
of the United States Government.

1/ Work done in part under contract Nonr 2582(00), Task
NR 042-200 of the Office of Naval Research.

CONTRIBUTIONS TO THE THEORY OF RANK ORDER STATISTICS:

Computation Rules for Probabilities of Rank Orders

I. Richard Savage

1. Introduction.

For most distributions the computations of the probabilities of rank orders (non null case) involve either difficult multiple integrations or extensive Monte Carlo sampling [1,2,3]. In this note rules are given for computing the probabilities of rank orders for the one and two sample problems [1,2]. For the one-sample problem the rule permits the computations for samples of size n from the results with samples of $n+1$. For the two-sample problem the rule permits the computations for samples of size m and n from the results with samples of $m+1$ and n (m and $n+1$). Since most computations done analytically are built up from smaller to larger sample sizes these results will, for that case, have limited value, e.g., in checking numerical work. For Monte Carlo sampling, however, there is no reason for starting with the smaller samples and in this case the rules will be of service.

2. One-Sample Rule.

Let $P_n(z)$ be the probability of the rank order $z = (z_1, \dots, z_n)$ where $z_i = 0$ (1) if the i -th smallest of the observed absolute values was from a negative (positive) observed value, e.g., if the observed values are (2.2, -.7, .5, -1.1, 3.0) then $z = (10011)$.

Rule I. To compute $P_n(z)$ add all $[2(n+1)$ in number] the $P_{n+1}(z^{ij})$ and divide by $(n+1)$ where

$$z^{ij} = (z_1, \dots, i, z_j, \dots, z_n) \quad i = 0, 1 \text{ and } j = 1, \dots, n+1.$$

Note a. z_{n+1} is undefined and actually is not used.

Note b. Several of the z^{ij} will be the same.

Note c. The rule can be obtained using the analytic expressions for $P_n(z)$ given in [2]. Another proof can be obtained from noting that after the sample of size n is formed an additional observation must fall either between existing observations or before them or after them.

Example I. Numerical results for the one-sample problem are not available. The following, however, suggests the kind of computing formulae that could be used. For $n=3$,

$$P_3(010) = [P_4(1010) + P_4(0010) + P_4(0110) + P_4(0010) + P_4(0110) + P_4(0100) + P_4(0101) + P_4(0100)]/4.$$

3. Two-Sample Rule.

Let $P_{m,n}(z)$ be the probability of the rank order $z = (z_1, \dots, z_{m+n})$ where $z_i = 0$ (1) if the i -th smallest of the observed values was from the first (second) sample, e.g., if the observed values in the first sample were $(-1.5, 2.6)$, in the second sample $(3.4, -.9)$ then $z = (0101)$.

Rule II. To compute $P_{m,n}(z)$ add all $[(m+n+1) \text{ in number}]$ of the $P_{m+1,n}(z^j)$ and divide by $(m+1)$ where

$$z^j = (z_1, \dots, 0, z_j, \dots, z_{m+1}) \quad j=1, \dots, (m+n+1).$$

Note a. z_{m+n+1} is undefined but is not actually used.

Note b. Several of the z^j will be the same.

Note c. The roles of m and n can be interchanged in the obvious manner.

Note d. The rule can be obtained using the analytic expression for $P_{m,n}(z)$ given in [1]. Another proof can be obtained by noting that after the samples of size m and n have been obtained an additional observation from the first

population must either be between a pair of the observations of the original $m+n$ or before or after them.

Example II. For the two-sample problem with $m=3$ and $n=2$,

$$P_{3,2}(00011)=[P_{3,3}(100011)+P_{3,3}(010011)+P_{3,3}(001011)+3P_{3,3}(000111)]/3.$$

Teichroew [3] gives .0394 as the exact value, and .0410 as the Monte Carlo value (2000 samples) when the two populations are normal with means differing by $1/2$ of the common standard deviation. Using Teichroew's [3] Monte Carlo results for $m=3$, $n=3$ (4000 samples) in the above formula one obtains

$$P_{3,2}(00011)=[.03250+.01825+.011875+3(.01675)]/3=.03992.$$

Additional results for $m=3$, $n=2$ could be obtained from $m=4$, $n=2$ and from $m=4$, $n=3$ via $m=3$, $n=3$ [3].

References

- [1] I. R. Savage, "Contributions to the theory of rank order statistics-- the two-sample case," Ann. of Math. Stat., Vol. 27 (1956), pp. 590-615.
- [2] I. R. Savage, "Contributions to the theory of rank order statistics-- the one-sample case,"--to be published.
- [3] D. Teichroew, "Empirical power functions for nonparametric two-sample tests for small samples," Ann. of Math. Stat., Vol. 26 (1955), pp. 340-344.

Department of Statistics
University of Minnesota

BASIC DISTRIBUTION LIST FOR UNCLASSIFIED TECHNICAL REPORTS

for contract Nonr 2582(00), Task NR 042-200

<u>ADDRESS</u>		<u>ADDRESS</u>	
Head, Statistics Branch Office of Naval Research Washington 25, D. C.	2	Professor W.G. Cochran Department of Statistics Harvard University Cambridge, Massachusetts	1
Commanding Officer Office of Naval Research Branch Office Navy Nr. 100 Fleet Post Office New York, New York	2	Professor Benjamin Epstein Applied Mathematics and Statistics Lab. Stanford University Stanford, California	1
ASTIA Document Service Center Arlington Hall Station Arlington 12, Virginia	10	Professor W. Hirsch Institute of Mathematical Sciences New York University New York 3, New York	1
Office of Technical Services Department of Commerce Washington 25, D. C.	1	Dr. Paul G. Hoel Department of Mathematics University of California Los Angeles 24, California	1
Technical Information Officer Naval Research Laboratory Washington 25, D. C.	6	Professor Harold Hotelling Associate Director Institute of Statistics University of North Carolina Chapel Hill, North Carolina	1
Professor T. W. Anderson Department of Mathematical Statistics Columbia University New York 27, New York	1	Professor L. Hurwicz School of Business Administration University of Minnesota Minneapolis 14, Minnesota	1
Professor Z. W. Birnbaum Laboratory of Statistical Research Department of Mathematics University of Washington Seattle 5, Washington	1	Professor Leo Katz Department of Statistics Michigan State University East Lansing, Michigan	1
Professor A. H. Bowker Applied Mathematics and Statistics Lab. Stanford University Stanford, California	1	Professor Oscar Kempthorne Statistics Laboratory Iowa State College Ames, Iowa	1
Professor Ralph A. Bradley Department of Statistics and Statistical Laboratory Virginia Polytechnic Institute Blacksburg, Virginia	1	Dr. Carl F. Kossack Statistical Laboratory Engineering Administration Building Purdue University Lafayette, Indiana	1
Professor Herman Chernoff Applied Mathematics and Statistics Lab. Stanford University Stanford, California	1		

ADDRESS

Professor Gerald J. Lieberman
Applied Mathematics and Statistics
Laboratory
Stanford University
Stanford, California 1

Professor William G. Madow
Department of Statistics
Stanford University
Stanford, California 1

Professor J. Neyman
Department of Statistics
University of California
Berkeley 4, California 1

Professor Herbert Robbins
Mathematical Statistics Department
Columbia University
New York 27, New York 1

Professor Murray Rosenblatt
Department of Mathematics
Indiana University
Bloomington, Indiana 1

Professor L. J. Savage
Statistical Research Laboratory
Chicago University
Chicago 37, Illinois 1

Dr. Herbert Solomon
Teachers College
Department of Statistics
Columbia University
New York 27, New York 1

Professor Frank Spitzer
Department of Mathematics
University of Minnesota
Minneapolis 14, Minnesota 1

Professor S. S. Wilks
Department of Mathematics
Princeton University
Princeton, New Jersey 1

ADDRESS

Professor Evan J. Williams
Institute of Statistics
State College Section
North Carolina State College
Raleigh, North Carolina 1

Professor J. Wolfowitz
Department of Mathematics
Cornell University
Ithaca, New York 1